ISCR

Institute for Scientific Computing Research

Laboratory Directed Research and Development Project Research Summaries

LDRD Projects

The Laboratory Directed Research and Development Program (LDRD) is one of LLNL's most important vehicles for developing and extending the Laboratory's intellectual foundations and maintaining its vitality as a premier research institution. The program selects and funds an investment portfolio of high-risk, high-potential-payoff R&D projects that foster the development of new scientific and technical capabilities in support of the Laboratory's evolving national security missions. The Computation Directorate (through the ISCR) and the University Relations Program jointly oversee a portion of the LDRD program in Exploratory Research in the Institutes (ERI) that focuses on computing technologies. ERIs have the specific charter of engaging academia as a part of their research activities. This section contains the annual reports for the four ERI LDRD projects supported in FY 2003.

Parallel Graph Algorithms for Complex Networks

Edmond Chow

Center for Applied Scientific Computing-LLNL

ISCR Institute for Scientific Computing Research

Summary

Many scientific, technological, and social networks—such as metabolic networks, the World-Wide Web, and human interaction graphs —are being studied as "complex networks": graph topologies that appear random, but have very specific clustering and vertex-degree properties. At LLNL, a particular complex network called a "semantic graph" is being developed to organize and search information across diverse and extremely large databases. Such graph data structures will contain vast amounts of information and must be stored and queried on distributed-memory parallel computers. Our objective is to develop algorithms and research software that exploit the properties of complex networks and enable fast and efficient use of semantic graphs on parallel computers.

In FY03, we began developing a software infrastructure for semantic graphs. In particular, we have developed generators for ontologies and for semantic graphs that are consistent with these ontologies. We have developed heuristic algorithms that use ontologies to accelerate point-to-point searches in semantic graphs. We have developed a preliminary algorithm for parallelizing these searches and are beginning to understand what objectives must be optimized in partitioning semantic graphs. Further, we are applying linear algebraic techniques for clustering vertices in graphs, based on their topology. Finally, we are investigating a partitioning approach based on classifying relationships as long-range and short-range, assuming that there is a spatial embedding of the semantic graph.

In FY04, we will develop partitioning algorithms that utilize augmented ontologies, i.e., ontologies that are annotated with the frequencies of edge and vertex types in the semantic graph. The simplest approach is to partition the augmented ontology as a weighted graph and induce this partitioning on the semantic graph. However, we expect better results to be achieved if the augmented ontology is refined to have some of the major topological properties of the semantic graph. This refinement can be accomplished by clustering similar vertices or edges. In FY04, we will continue to pursue promising approaches discovered in the previous year. These include dynamic partitions based on analysts' queries. Finally, in FY04, we will begin the parallel implementation of some of our algorithms.

This work has relevance to a number of different current and future applications at LLNL. As one example, the Department of Homeland Security has many different organizations, each with its own extremely large databases of information. This technology of semantic graphs with fast and efficient search algorithms could provide a method for working with all of the data as one enormous database.

Enabling Large-Scale Data Access

Terence Critchlow

Center for Applied Scientific Computing-LLNL

Summary

This project's goal is to develop an infrastructure capable of providing scientists with access to large numbers of data sources through a single, intuitive interface. This interface will simplify scientists' interaction with data and enable them to answer more complex questions than currently possible. Our infrastructure makes extensive use of a novel metadata infrastructure to identify and describe Web-based interfaces to complex, scientific data sources. The resulting descriptions are then passed to an extended version of XWrap, a wrapper-generation program. We have developed an initial prototype of this infrastructure and have performed initial testing of the system.

This project will position LLNL as a leader in several technology areas, including data integration, bioinformatics, metadata, Web-aware agents, and wrapper generators by developing an infrastructure for accessing extensive data sources through a single interface. We plan to demonstrate the feasibility of this approach by automatically identifying sequence-similarity search interfaces across a variety of genomics Web sites.

We met all of our deliverables for this fiscal year. Specifically, we wrote the initial specification of the service-class description format, defined a new version of the interface-description format, completed the initial version of the Web spider and used it to successfully identify 17 of 25 randomly selected BLAST sites, extended XWrap Composer to accept complex interface description as input and to generate wrappers based on these descriptions, used Composer to generate wrappers for two complex BLAST interfaces, extended the DataFoundry interface to interact with BLAST wrappers, and incorporated two of the wrappers into this infrastructure.

Our milestones for FY04 are to (1) implement complex interface-identification capabilities that will support automatic identification of simple indirection pages, (2) finalize interface-description and service-class-description formats, (3) extend the Web spider to support automatic generation of interface descriptions, (4) refine Composer to accept the final interface-description format and use it to generate complex wrappers, and (5) demonstrate end-to-end automatic wrapper-interface generation capabilities for both simple and complex interfaces.

This work supports national security and other LLNL missions by benefiting ongoing programs at LLNL such as nonproliferation and stockpile stewardship, which could better utilize information from a wide variety of sources if this project is successful. Our infrastructure would simplify creating an interface that combines local data with related information publicly available over the Internet, such as scientific publications, chemistry databases, urban planning information, and census data.

Robust Real-Time Techniques for Detection and Tracking in Video

Chandrika Kamath

Center for Applied Scientific Computing-LLNL

ISCR Institute for Scientific Computing Research

Summary

The use of video cameras for monitoring and surveillance is becoming prevalent in diverse applications such as battlefield awareness through the identification of troop movement, national security, monitoring of traffic, detection of pedestrians, identification of anomalous behavior in a parking lot or near an ATM, etc. While a single image provides a snapshot of a scene, the different frames of a video taken over time register the dynamics, making it possible to capture motion.

The two key tasks in mining video data are the detection of moving objects and the tracking of these objects as they move over time. Once the objects have been detected and tracked, this information can be used to build models of normal behavior in the scene, enabling us to flag behavior that is abnormal. In this Laboratory Directed Research and Development (LDRD) project, we are investigating detection and tracking algorithms that are both accurate and robust (to minimize false alarms and missed positives) and real-time (to allow for corrective action). In particular, we are interested in video taken by a stationary camera under adverse conditions such as fog, or at a low resolution, or at a low frame rate. Our algorithms and software include techniques to separate the moving foreground from the background, to extract features representing the foreground objects, and to track these objects from frame to frame, followed by post-processing to smooth the tracks.

Since the start of the project in March 2003, we have identified realistic videos of traffic intersections taken under different conditions, designed a software infrastructure to handle color video, implemented several techniques for background subtraction, and evaluated their performance on our test videos. The best performance was obtained by using a mixture of Gaussians to maintain the background; however, this technique is quite expensive. A simple median of the frames also worked well and computationally was much less expensive. Further reduction in compute time was obtained by using an approximation to the median. In addition, we implemented simple schemes to calculate features such as the velocity of the moving objects.

We also conducted research on block matching techniques and found that it was effective for moderate resolution video. We could improve the performance in the presence of small camera motion by keeping a frame history. Though our main focus has been on tracking vehicles, we also supported a collaboration with the University of Colorado-Boulder on the tracking of multiple people.

In FY04, we will develop more accurate schemes to extract features such as the velocity and address the problem of tracking the objects across frames. We will also investigate how the techniques carry over to the lower resolution, lower frame rate videos.

This capability to detect and track moving objects in video supports LLNL's national security mission. For instance, this capability would be useful in counter-terrorist applications.

ViSUS: Visualization Streams for Ultimate Scalability

Valerio Pascucci

Center for Applied Scientific Computing (CASC)-LLNL

Summary

Modern scientific simulations and experimental settings produce ever-increasingly large amounts of data that traditional tools are not able to visualize in real time, especially on regular desktop computers. This inability of the scientists to interactively explore their data sets creates a frustrating slowdown in the overall process of scientific discovery.

The ViSUS project is developing a suite of progressive visualization algorithms and a data-streaming infrastructure to enable interactive exploration of large scientific datasets. The methodology optimizes the data flow in a pipeline of processing modules. Each module reads in input and writes in output a multi-resolution representation of a geometric model. This provides the flexibility to trade speed for accuracy, as needed. The dataflow is streamlined with progressive algorithms that map local geometric updates of the input into immediate updates of the output. A prototype streaming infrastructure will demonstrate the flexibility and scalability of this approach visualizing large data on a single desktop computer, on a cluster of personal computers, and on heterogeneous computing resources.

In the first two years the ViSUS project has published 19 peer-reviewed papers and produced software tools for the visualization of 2D/3D scalar fields and surface meshes. The main achievements for FY03 include:

- Completed an end-to-end prototype of the streaming infrastructure for the progressive monitoring of a Computational Fluid Dynamics (CFD) simulation (GP code).

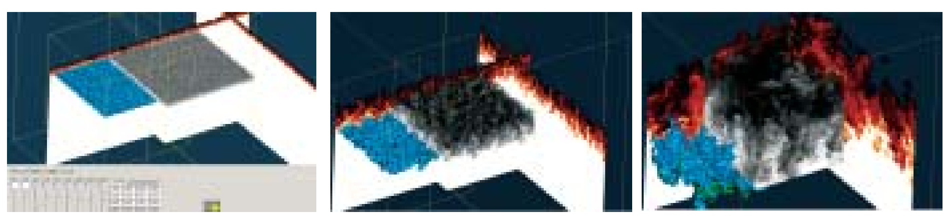

Fig. 1. Three timesteps of a Raleigh-Taylor simulation computed by the MIRANDA code (LLNL AX division) and visualized with the ViSUS Progressive Viewer. Images generated on a DELL laptop computer.

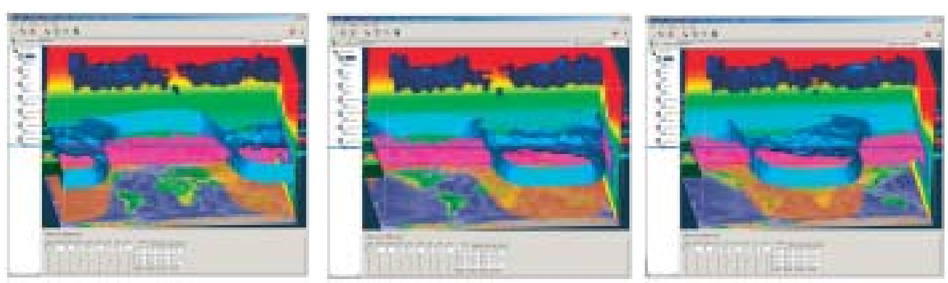

Fig. 2. Three timesteps of a climate modeling simulation computed by the IMPACT code (LLNL NCAR division) and visualized with the ViSUS Progressive Viewer. Images generated on a DELL laptop computer.

Continued

Summary continued

- Produced an initial prototype for the integration with the MIRANDA and IMPACT codes (see Figures 1 and 2).
- Developed an external memory technique allowing interactive traversal of multi-resolution surface meshes of arbitrary size.
- Completed the definition and prototype implementation of a wavelet model for our volumetric subdivision scheme, which will be used to extend our approach to non-rectilinear meshes.
- Completed a prototype implementation of an image cache engine that allows maintaining interactive data exploration even for slow rendering algorithms.

In FY04 we will bring the ViSUS technology to a level maturity and robustness that allows deployment to a few targeted users and demonstrates the practical advantages provided by our tools. The main milestones include:

- Develop new techniques that accelerate the isosurface construction process with occlusion culling, graphics hardware, and simple view-dependent adaptive refinements.
- Develop a new approach that allows for multi-resolution streaming of large triangulated meshes.
- Incorporate in the ViSUS Progressive Viewer the full slicing, isocontouring and volume rendering capabilities.
- Deploy the ViSUS streaming technology to the MIRANDA code (AX division) and IMPACT code (NCAR division).
- Provide a prototype viewer with new data-structures capable of handling the unstructured data format used in the HYDRA (National Ignition Facility [NIF]) simulation code;
- Provide easy installation procedures and lab-wide access to our software.

Use of our innovative, high-performance visualization techniques will allow interactive display of very large data sets on simple desktop workstations and the monitoring (or steering) of large parallel simulations. This will have specific applications to several of the DOE's and LLNL's missions, including stockpile stewardship, energy and environment, nonproliferation, biology, and basic science that use large-scale modeling and simulations. More specifically, the Laboratory will benefit from the new ViSUS technology at least at two levels. At the deployment level, the improved efficiency in the use of hardware resources will reduce the cost of visualization-hardware infrastructures. At the scientific level, the developed technology will reduce the overall time required for the design, simulation, and visualization cycle. For instance, the ability to remotely monitor large and expensive simulations will save computing resources through early termination and restart of erroneous test simulations.